The 48 hour annotation server

Ross Overbeek
Felloship for Interpretation of Genomes

December 2006

Background

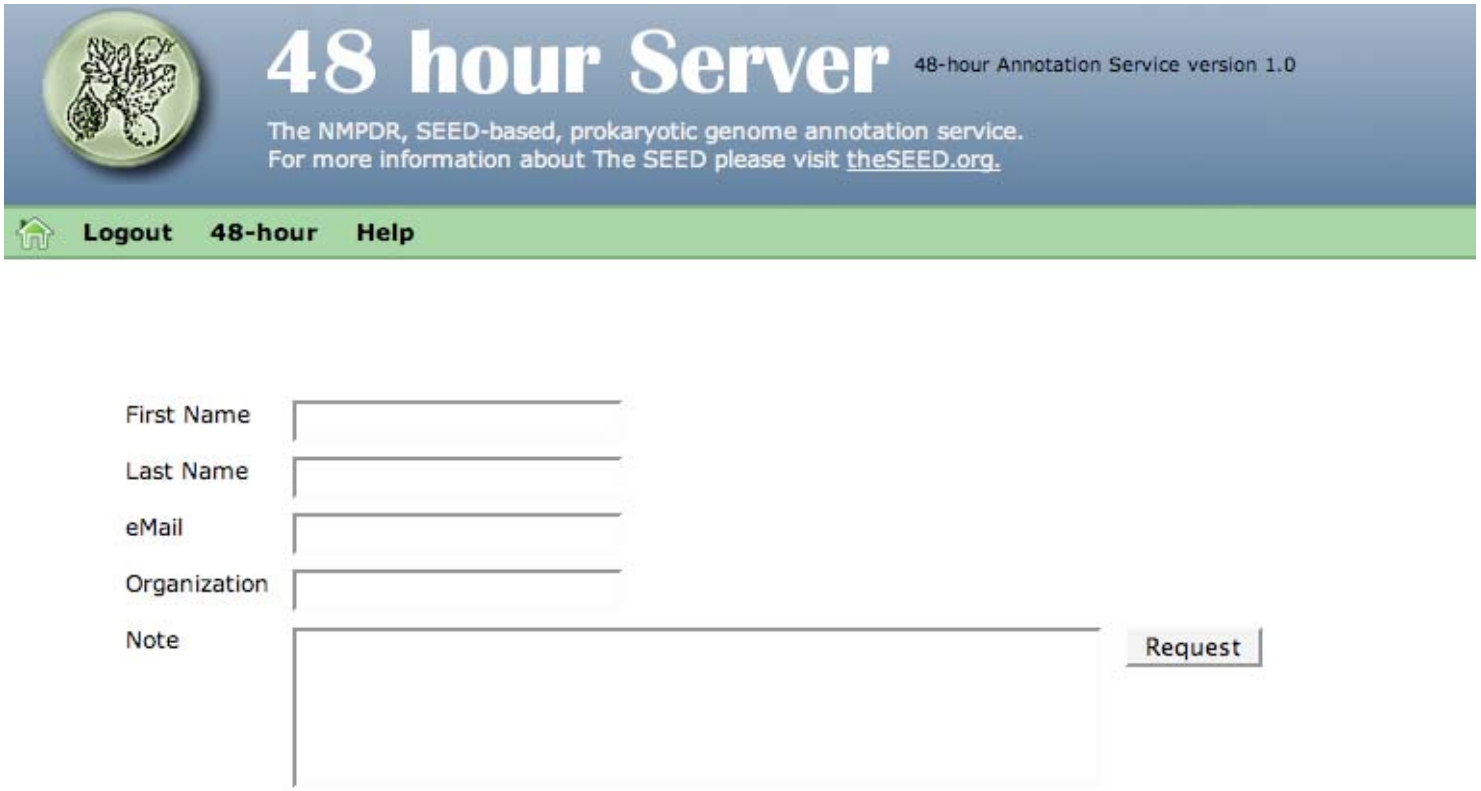
- Initially based on the type of analysis used only for “close-strain” annotation.
- **Goal**: to provide near-perfect annotation for close strains within our 5 pathogen families in an initial, rapid annotation
- **Insight**: with minor enhancements, the technology works well for all prokaryotic genomes!
- ⇨We provide a service where users can upload their genome and get a complete annotation in a 48 hour timeframe

Procedure

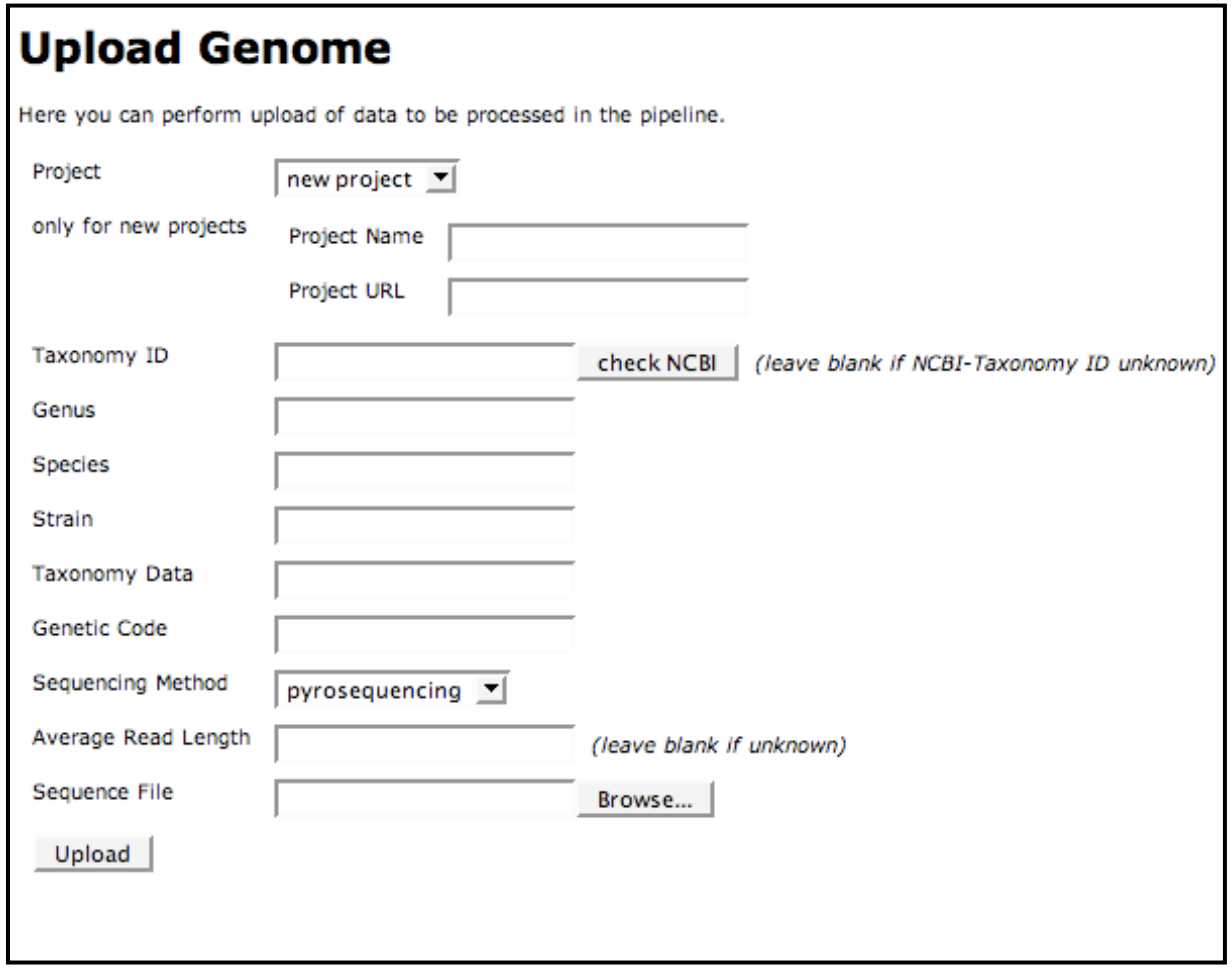
1. Register (user)
2. Submit genome (user)
3. Genome annotation process (server)
4. Quality controls (server)
5. Evaluation / Viewing (user)
6. Download results (user)
7. Delete the genome from the server (user)

Register for an account

- http://bioseed.mcs.anl.gov/48-hour/FIG/index48.cgi



Submit a genome



Upload Genome

Here you can perform upload of data to be processed in the pipeline.

Project: new project

only for new projects — Project Name: ; Project URL:

Taxonomy ID: [check NCBI] *(leave blank if NCBI-Taxonomy ID unknown)*

Genus:

Species:

Strain:

Taxonomy Data:

Genetic Code:

Sequencing Method: pyrosequencing

Average Read Length: *(leave blank if unknown)*

Sequence File: [Browse...]

[Upload]

- Fill in key pieces of information
- Provided information will be checked against input data

Genome annotation process

- Automated process consisting of:
  - Gene calling
  - Initial annotation of function
  - Initial metabolic reconstruction
- Process takes 1-7 hours depending on size and complexity of the genome
- ~20 genomes per day

⇨We call it 48 hour server to leave ourselves some breathing room

Quality control

Job Details #87

» Back to the Jobs Overview

Genome Upload has been successfully completed.

Genome:	234267.8 - Solibacter usitatus Ellin6076
Job:	#87
User:	batch
Date:	Mon Dec 4 17:16:27 2006

Rapid Propagation is currently in progress.

Quality Check has not yet started.

Quality Revision has not yet started.

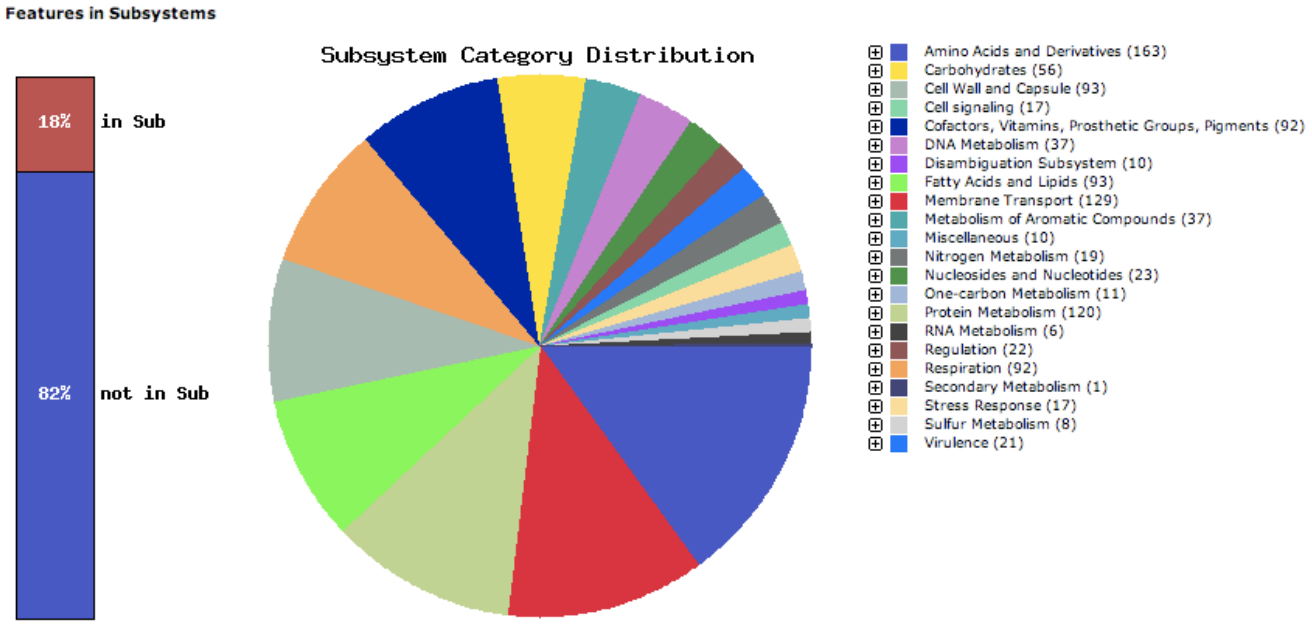
Similarity Computation has not yet started.

Auto Assignment has not yet started.

Evaluation / Viewing

Organism Overview

Genome	Desulfitobacterium hafniense Y51 (Taxonomy ID: 1381
Domain	Bacteria
Size	5,727,534 bp
Number of Contigs	1
Number of Subsystems	162
Number of Coding Sequences	5214
Number of RNAs	77



Evaluation / Viewing II

Genome Browser

Genome	Desulfitobacterium hafniense Y51 (Taxonomy ID: 138119)
Domain	Bacteria
Size	5,727,534 bp
Number of Contigs	1
Number of Subsystems	162
Number of Coding Sequences	5214
Number of RNAs	77

Protein Encoding Genes

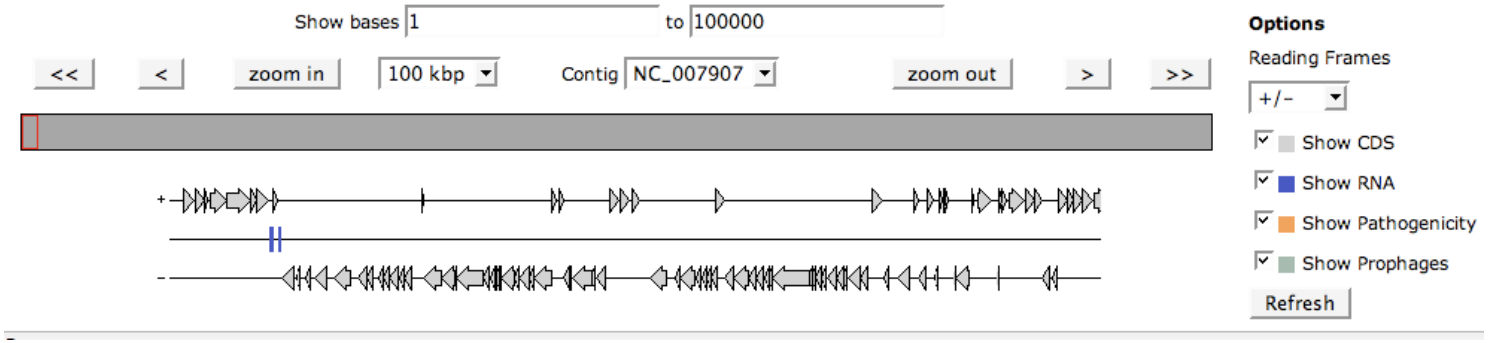
non-hypothetical	3726	71.46%
hypothetical	1488	28.54%

Browsing Genome Desulfitobacterium hafniense Y51

The genome browser offers a two frame view of the currently selected organism. If the genome is not completely assembled, the first contig will be selected.

On the middle line you will see features not associated to a reading frame, such as rnas, pathogenicity islands or prophages. You can change the number of displayed frames, choose which kinds of features are to be displayed, as well as change the zoom level of the display.

Hover over an item for a detailed description. **Clicking** an item will allow you to access its annotation overview.



Download results

- We provide the following formats:
  - Genbank
  - All formats supported by BioPerl (soon)
- Genomes can be deleted by the user at any time
- Genomes can be directly imported into the SEED if the user wishes

Conclusion

- Service for annotation of complete prokaryotic genomes available over the web
- Now operational - we have processed 50 genomes in the last two days
- It produces annotations somewhat worse than our best manual annotations (which are in subsystems)
- Addressing issues related to:
  - Frameshifts
  - Pseudo-genes
  - Prophages
  - Transposable elements
  - Start calls